Lecture 2:

Uncertainty, Probability and Information

Dr. Roman V Belavkin

Middlesex University

BIS4435

Historical background

Sources of uncertainty

What is probability?

Joint Probability and Independence

Uncertainty and Information

Reducing the Uncertainty in Data, Random variables

Decision-Making under Uncertainty

HISTORICAL BACKGROUND

1654 Blaise Pascal and Pierre Fermat

HISTORICAL BACKGROUND

1654 Blaise Pascal and Pierre Fermat

1657 Christian Huygens publishes *On Ratiocination in Dice Games*

HISTORICAL BACKGROUND

1654 Blaise Pascal and Pierre Fermat

1657 Christian Huygens publishes *On Ratiocination in Dice Games*

1760 Thomas Bayes (conditional probability)

HISTORICAL BACKGROUND

1654 Blaise Pascal and Pierre Fermat

1657 Christian Huygens publishes *On Ratiocination in Dice Games*

1760 Thomas Bayes (conditional probability)

1812 Pierre–Simon Laplace

HISTORICAL BACKGROUND

1654 Blaise Pascal and Pierre Fermat

1657 Christian Huygens publishes *On Ratiocination in Dice Games*

1760 Thomas Bayes (conditional probability)

1812 Pierre–Simon Laplace

1933 Andrey Kolmogorov's axioms

HISTORICAL BACKGROUND

1654	Blaise Pascal and Pierre Fermat
1657	Christian Huygens publishes *On Ratiocination in Dice Games*
1760	Thomas Bayes (conditional probability)
1812	Pierre–Simon Laplace
1933	Andrey Kolmogorov's axioms
1920-1940	Ronald Fisher, Abraham Wald (statistics)

HISTORICAL BACKGROUND

1654	Blaise Pascal and Pierre Fermat
1657	Christian Huygens publishes *On Ratiocination in Dice Games*
1760	Thomas Bayes (conditional probability)
1812	Pierre–Simon Laplace
1933	Andrey Kolmogorov's axioms
1920-1940	Ronald Fisher, Abraham Wald (statistics)
1948	Claude Shannon (Information theory)

SOURCES OF UNCERTAINTY

Complexity : the number of possible states of a system in question can be too large (e.g. predict how a chess game can develop after 10 moves?)

SOURCES OF UNCERTAINTY

Complexity : the number of possible states of a system in question can be too large (e.g. predict how a chess game can develop after 10 moves?)

Ignorance : some important information about the system may not be available

SOURCES OF UNCERTAINTY

Complexity : the number of possible states of a system in question can be too large (e.g. predict how a chess game can develop after 10 moves?)

Ignorance : some important information about the system may not be available

Randomness : the system may be random by nature, and thus the uncertainty is irreducible.

WHAT IS PROBABILITY?

Definition

The uncertainty about some event E can range from impossible to certain. Let us denote **probability** of event E by $P(E)$ such that

$P(E) = 0$ means E is impossible;

$P(E) = 1$ means E is certain.

WHAT IS PROBABILITY?

Definition

The uncertainty about some event E can range from impossible to certain. Let us denote **probability** of event E by $P(E)$ such that

$P(E) = 0$ means E is impossible;

$P(E) = 1$ means E is certain.

Thus, probability is a number between 0 and 1.

$$\text{(Impossible)} \qquad 0 \leq P(E) \leq 1 \qquad \text{(Certain)}$$

WHAT IS PROBABILITY?

Definition

The uncertainty about some event E can range from impossible to certain. Let us denote **probability** of event E by $P(E)$ such that

$P(E) = 0$ means E is impossible;

$P(E) = 1$ means E is certain.

Thus, probability is a number between 0 and 1.

$$\text{(Impossible)} \qquad 0 \leq P(E) \leq 1 \qquad \text{(Certain)}$$

Example

For a fair coin, $P(heads) = \frac{1}{2} = 0.5$

ADDITIVITY OF PROBABILITIES

- It is certain that at least one of the alternative events will happen.

ADDITIVITY OF PROBABILITIES

- It is certain that at least one of the alternative events will happen.
- If E_1, E_2, ..., E_n are n alternative (disjoint) events, then the fact that at least one of them will certainly happen can be written as

$$P(E_1) + P(E_2) + \cdots + P(E_n) = 1$$

ADDITIVITY OF PROBABILITIES

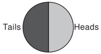
- It is certain that at least one of the alternative events will happen.
- If E_1, E_2, ..., E_n are n alternative (disjoint) events, then the fact that at least one of them will certainly happen can be written as

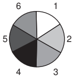
$$P(E_1) + P(E_2) + \cdots + P(E_n) = 1$$

Example

For a fair coin and a fair dice we have

$$\frac{1}{2} + \frac{1}{2} = 1 \qquad\qquad \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$





WHERE DO PROBABILITIES COME FROM?

- If there are n disjoint events, then we could assume that all

$$P(E_1) = P(E_2) = \cdots = P(E_n) = \frac{1}{n}$$

WHERE DO PROBABILITIES COME FROM?

- If there are n disjoint events, then we could assume that all

$$P(E_1) = P(E_2) = \cdots = P(E_n) = \frac{1}{n}$$

- It would be much better to use the empirical frequency function

$$P(E_i) \approx \frac{n(E_i)}{n} = \frac{\text{no. of times event } E_i \text{ occurs}}{\text{no. of independent tests}}$$

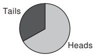
WHERE DO PROBABILITIES COME FROM?

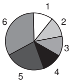
- If there are n disjoint events, then we could assume that all

$$P(E_1) = P(E_2) = \cdots = P(E_n) = \frac{1}{n}$$

- It would be much better to use the empirical frequency function

$$P(E_i) \approx \frac{n(E_i)}{n} = \frac{\text{no. of times event } E_i \text{ occurs}}{\text{no. of independent tests}}$$





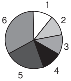
WHERE DO PROBABILITIES COME FROM?

- If there are n disjoint events, then we could assume that all

$$P(E_1) = P(E_2) = \cdots = P(E_n) = \frac{1}{n}$$

- It would be much better to use the empirical frequency function

$$P(E_i) \approx \frac{n(E_i)}{n} = \frac{\text{no. of times event } E_i \text{ occurs}}{\text{no. of independent tests}}$$

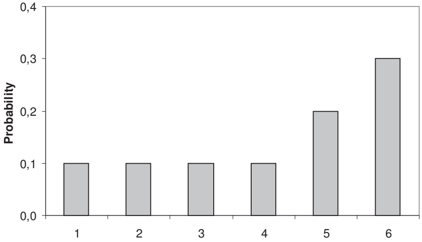


Example

Flip a coin or roll a dice several times to estimate the probabilities.

PROBABILITY DISTRIBUTIONS

We can plot probabilities of all events on a graph, which shows probability **distribution**



JOINT PROBABILITY

- What is the probability of co–occurrence of several events? (e.g. clouds and rain)

JOINT PROBABILITY

- What is the probability of co–occurrence of several events? (e.g. clouds and rain)
- Do these events occur together simply by chance?

JOINT PROBABILITY

- What is the probability of co–occurrence of several events? (e.g. clouds and rain)
- Do these events occur together simply by chance?
- Probability $P(E_1, E_2)$ is called joint probability of E_1 and E_2.

E_1	E_2
heads	heads
heads	tails
tails	heads
tails	tails

CONDITIONAL PROBABILITY and INDEPENDENCE

- $P(E_1 \mid E_2)$ is the probability of E_1 conditional to event E_2 (i.e. if E_2 has happened).

CONDITIONAL PROBABILITY and INDEPENDENCE

- $P(E_1 \mid E_2)$ is the probability of E_1 conditional to event E_2 (i.e. if E_2 has happened).
- Joint probability is $P(E_1, E_2) = P(E_1 \mid E_2)P(E_2)$ and

$$P(E_1 \mid E_2) = \frac{P(E_1, E_2)}{P(E_2)}$$

CONDITIONAL PROBABILITY and INDEPENDENCE

- $P(E_1 \mid E_2)$ is the probability of E_1 conditional to event E_2 (i.e. if E_2 has happened).
- Joint probability is $P(E_1, E_2) = P(E_1 \mid E_2)P(E_2)$ and

$$P(E_1 \mid E_2) = \frac{P(E_1, E_2)}{P(E_2)}$$

- If E_1 is independent of E_2, then $P(E_1 \mid E_2) = P(E_1)$ and

$$P(E_1, E_2) = P(E_1)P(E_2)$$

SIMPLE CREDIT SCORE EXAMPLE

- You selected at random records of 20 customers and divided them based on *homeowner* and *credit score*.

		Credit score	
		Low	High
Homeowner	No	7	3
	Yes	2	8

SIMPLE CREDIT SCORE EXAMPLE

- You selected at random records of 20 customers and divided them based on *homeowner* and *credit score*.

		Credit score	
		Low	High
Homeowner	No	7	3
	Yes	2	8

- Can you tell from this data whether credit score and homeownership depend on each other?

SIMPLE CREDIT SCORE EXAMPLE (sol.)

		Credit score		
		Low	High	
Homeowner	No	7	3	**10**
	Yes	2	8	**10**
		9	**11**	**20**

SIMPLE CREDIT SCORE EXAMPLE (sol.)

		Credit score		
		Low	High	
Homeowner	No	7	3	**10**
	Yes	2	8	**10**
		9	**11**	**20**

$$P(\mathrm{H}, \mathrm{C}) = \left\{ \frac{7}{20}, \frac{2}{20}, \frac{3}{20}, \frac{8}{20} \right\}$$

SIMPLE CREDIT SCORE EXAMPLE (sol.)

		Credit score		
		Low	High	
Homeowner	No	7	3	**10**
	Yes	2	8	**10**
		9	**11**	**20**

$$P(\mathrm{H}, \mathrm{C}) = \left\{ \frac{7}{20}, \frac{2}{20}, \frac{3}{20}, \frac{8}{20} \right\}$$

$$P(\mathrm{H}) = \left\{ \frac{10}{20}, \frac{10}{20} \right\}, \qquad P(\mathrm{C}) = \left\{ \frac{9}{20}, \frac{11}{20} \right\}$$

SIMPLE CREDIT SCORE EXAMPLE (sol.)

		Credit score		
		Low	High	
Homeowner	No	7	3	**10**
	Yes	2	8	**10**
		9	**11**	**20**

$$P(\mathrm{H},\mathrm{C}) = \left\{\frac{7}{20},\frac{2}{20},\frac{3}{20},\frac{8}{20}\right\}$$

$$P(\mathrm{H}) = \left\{\frac{10}{20},\frac{10}{20}\right\}, \qquad P(\mathrm{C}) = \left\{\frac{9}{20},\frac{11}{20}\right\}$$

Our simple test for independence $P(\mathrm{H},\mathrm{C}) \stackrel{?}{=} P(\mathrm{H})P(\mathrm{C})$

UNCERTAINTY and INFORMATION

- If each event can have 2 states (i.e. True, False), then for a system that consists of H number of events there are

$$M = 2^H \text{ possible states}$$

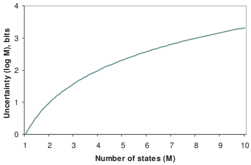
$$\text{Information} = \text{Uncertainty before} - \text{Uncertainty after}$$

UNCERTAINTY and INFORMATION

- If each event can have 2 states (i.e. True, False), then for a system that consists of H number of events there are

$$M = 2^H \text{ possible states}$$

- To measure uncertainty we can use $H = \log_2 M$ (aka *entropy*)



$$\text{Information} = \text{Uncertainty before} - \text{Uncertainty after}$$

INFORMATION and PROBABILITY

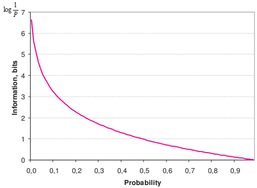
- What if probabilities are not equal? (i.e. $P(E) \neq \frac{1}{M}$)

INFORMATION and PROBABILITY

- What if probabilities are not equal? (i.e. $P(E) \neq \frac{1}{M}$)
- We can express the uncertainty as $H = \log_2 \frac{1}{P(E)}$

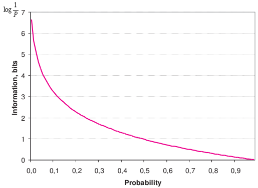
INFORMATION and PROBABILITY

- What if probabilities are not equal? (i.e. $P(E) \neq \frac{1}{M}$)
- We can express the uncertainty as $H = \log_2 \frac{1}{P(E)}$



INFORMATION and PROBABILITY

- What if probabilities are not equal? (i.e. $P(E) \neq \frac{1}{M}$)
- We can express the uncertainty as $H = \log_2 \frac{1}{P(E)}$



Example

Compare information from observing events with probabilities $\frac{1}{8}$ and $\frac{1}{2}$

REDUCING THE UNCERTAINTY IN DATA

Case:	Age	Gender	M. Income ($ K)	M. Expenses ($ K)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

REDUCING THE UNCERTAINTY IN DATA

Case:	Age	Gender	M. Income ($ K)	M. Expenses ($ K)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

Variables Age = [1,2,...,100], Gender =[0 (Female), 1 (Male)]

REDUCING THE UNCERTAINTY IN DATA

Case:	Age	Gender	M. Income ($ K)	M. Expenses ($ K)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

Variables Age = [1,2,...,100], Gender =[0 (Female), 1 (Male)]

How often does each value appear in the data?

REDUCING THE UNCERTAINTY IN DATA

Case:	Age	Gender	M. Income ($ K)	M. Expenses ($ K)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

Variables Age = [1,2,...,100], Gender =[0 (Female), 1 (Male)]

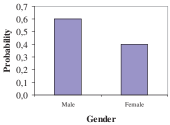
How often does each value appear in the data?

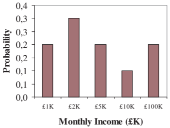
Random variables if each value is associated with probability

$$P(\text{Male}) = \frac{3}{5}, \quad P(\text{Female}) = \frac{2}{5}, \quad (P(\text{Male}) + P(\text{Female}) = 1)$$

IS THERE STRUCTURE IN DATA?

- Using the concept of random variables, we can analyse the distributions of each variable in the database

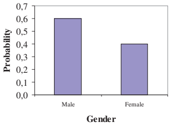


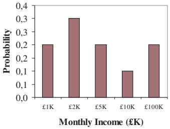


- Each case in the database can be seen as a complex (joint) event (e.g. Case 1 is Age=21, Gender=Female, etc).

IS THERE STRUCTURE IN DATA?

- Using the concept of random variables, we can analyse the distributions of each variable in the database



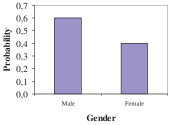


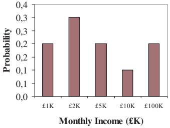
- Each case in the database can be seen as a complex (joint) event (e.g. Case 1 is Age=21, Gender=Female, etc).
- Thus, the whole database can be seen as a joint probability

$$P(\text{Case}) = P(\text{Age, Gender, Income, Expenses, H. owner, C. score})$$

IS THERE STRUCTURE IN DATA?

- Using the concept of random variables, we can analyse the distributions of each variable in the database





- Each case in the database can be seen as a complex (joint) event (e.g. Case 1 is Age=21, Gender=Female, etc).
- Thus, the whole database can be seen as a joint probability

$$P(\text{Case}) = P(\text{Age}, \text{Gender}, \text{Income}, \text{Expenses}, \text{H. owner}, \text{C. score})$$

- Are these variables independent or not?

MEASURES of LOCATION

- Answer questions such as '*What is the most probable value?*', '*What value should I expect in the long term?*'

MEASURES of LOCATION

- Answer questions such as '*What is the most probable value?*', '*What value should I expect in the long term?*'
- If variable x has n possible values X_1, X_2,...,X_n with probabilities $P(X_1)$, $P(X_2)$,...,$P(X_n)$, then we can compute the expected value

$$E\{x\} = X_1 P(X_1) + X_2 P(X_2) + \cdots + X_n P(X_n) = \sum_{i=1}^{n} X_i \, P(X_i)$$

MEASURES of LOCATION

- Answer questions such as '*What is the most probable value?*', '*What value should I expect in the long term?*'
- If variable x has n possible values X_1, X_2,...,X_n with probabilities $P(X_1)$, $P(X_2)$,...,$P(X_n)$, then we can compute the expected value

$$E\{x\} = X_1 P(X_1) + X_2 P(X_2) + \cdots + X_n P(X_n) = \sum_{i=1}^{n} X_i \, P(X_i)$$

- If all $P(x) = \frac{1}{n}$, then $E\{x\}$ is simply the average (the mean) value.

MEASURES of LOCATION (cont.)

Example

Variable Age has values 21, 18, 50, 23, 40. Each values occurs once, therefore $P(x) = \frac{1}{5}$ and the expected value is

MEASURES of LOCATION (cont.)

Example

Variable Age has values 21, 18, 50, 23, 40. Each values occurs once, therefore $P(x) = \frac{1}{5}$ and the expected value is

$$E\{\text{Age}\} = \frac{21 + 18 + 50 + 23 + 40}{5} = 30,4$$

which is also the average Age.

CENTRE OF GRAVITY

- What is the 'average' case?

CENTRE OF GRAVITY

- What is the 'average' case?
- The expected value for a joint distribution of m random variables x_1, x_2,...,x_m is a point in an m–dimensional space with coordinates given by expectations of each of the m variables, and is called the centre of gravity

$$E\{x\} = (E\{x_1\}, E\{x_2\}, \ldots, E\{x_m\})$$

CENTRE OF GRAVITY

- What is the ‘average’ case?
- The expected value for a joint distribution of m random variables x_1, x_2,...,x_m is a point in an m–dimensional space with coordinates given by expectations of each of the m variables, and is called the centre of gravity

$$E\{x\} = (E\{x_1\}, E\{x_2\}, \ldots, E\{x_m\})$$

- For our data, this is the expected case (i.e the average case)

$$E\{\text{Case}\} = (E\{\text{Age}\}, E\{\text{Gender}\}, \ldots, E\{\text{C. score}\})$$

MEASURES of DISPERSION

- Answer questions such as '*What is the range of the variable?*', '*What risk is associated with the variable?*'

MEASURES of DISPERSION

- Answer questions such as '*What is the range of the variable?*', '*What risk is associated with the variable?*'
- We can compute the average deviation from the expected value

$$E\{|x - E\{x\}|\} = \sum_{i=1}^{n} |X_i - E\{x\}| P(X_i)$$

MEASURES of DISPERSION

- Answer questions such as '*What is the range of the variable?*', '*What risk is associated with the variable?*'
- We can compute the average deviation from the expected value

$$E\{|x - E\{x\}|\} = \sum_{i=1}^{n} |X_i - E\{x\}| P(X_i)$$

- Or the average squared deviation, called the variance

$$Var\{x\} = E\{|x - E\{x\}|^2\} = \sum_{i=1}^{n} |X_i - E\{x\}|^2 P(X_i)$$

MEASURES of DISPERSION

- Answer questions such as '*What is the range of the variable?*', '*What risk is associated with the variable?*'
- We can compute the average deviation from the expected value

$$E\{|x - E\{x\}|\} = \sum_{i=1}^{n} |X_i - E\{x\}| P(X_i)$$

- Or the average squared deviation, called the variance

$$Var\{x\} = E\{|x - E\{x\}|^2\} = \sum_{i=1}^{n} |X_i - E\{x\}|^2 P(X_i)$$

- Standard deviation is $Sdev\{x\} = \sqrt{Var\{x\}}$

MEASURES of DISPERSION (cont.)

Example

Find $Var\{\text{Age}\}$ and $Sdev\{\text{Age}\}$?

1. Earlier we found $E\{\text{Age}\} = 30,4$.

MEASURES of DISPERSION (cont.)

Example

Find $Var\{\text{Age}\}$ and $Sdev\{\text{Age}\}$?

1. Earlier we found $E\{\text{Age}\} = 30,4$.
2. We need to find squared deviations from 30,4.

$$(21 - 30,4)^2\,,\ (18 - 30,4)^2\,,\ (50 - 30,4)^2\,,\ (23 - 30,4)^2\,,\ (40 - 30,4)^2$$

MEASURES of DISPERSION (cont.)

Example

Find $Var\{\text{Age}\}$ and $Sdev\{\text{Age}\}$?

1. Earlier we found $E\{\text{Age}\} = 30,4$.
2. We need to find squared deviations from 30,4.

$$(21-30,4)^2,\ (18-30,4)^2,\ (50-30,4)^2,\ (23-30,4)^2,\ (40-30,4)^2$$

3. Then we multiply each by $P(\text{Age}) = \frac{1}{5}$, and their sum gives the variance

$$Var\{\text{Age}\} = \frac{1}{5}((21-30,4)^2 + \cdots + (40-30,4)^2) = 154,64$$

MEASURES of DISPERSION (cont.)

Example

Find $Var\{\text{Age}\}$ and $Sdev\{\text{Age}\}$?

1. Earlier we found $E\{\text{Age}\} = 30,4$.
2. We need to find squared deviations from 30,4.

$$(21 - 30,4)^2,\ (18 - 30,4)^2,\ (50 - 30,4)^2,\ (23 - 30,4)^2,\ (40 - 30,4)^2$$

3. Then we multiply each by $P(\text{Age}) = \frac{1}{5}$, and their sum gives the variance

$$Var\{\text{Age}\} = \frac{1}{5}((21 - 30,4)^2 + \cdots + (40 - 30,4)^2) = 154,64$$

4. Standard deviation is a square root of the variance

$$Sdev(Age) = \sqrt{154,64} = 12,44$$

COVARIANCE

- Compares concentration of one variable with respect to another.

COVARIANCE

- Compares concentration of one variable with respect to another.
- If x and y are two random variables, then their covariance is

$$Cov(x, y) = E\{(x - E\{x\})(y - E\{y\})\}$$

COVARIANCE

- Compares concentration of one variable with respect to another.
- If x and y are two random variables, then their covariance is

$$Cov(x, y) = E\{(x - E\{x\})(y - E\{y\})\}$$

- Note that $Cov(x, y) = Cov(y, x)$ and $Cov(x, x) = Var\{x\}$

COVARIANCE

- Compares concentration of one variable with respect to another.
- If x and y are two random variables, then their covariance is

$$Cov(x, y) = E\{(x - E\{x\})(y - E\{y\})\}$$

- Note that $Cov(x, y) = Cov(y, x)$ and $Cov(x, x) = Var\{x\}$
- If x and y have 'similar' values, then $E\{x\} \approx E\{y\}$ and

$$Cov(x, y) \approx Var\{x\} \approx Var\{y\}$$

COVARIANCE

- Compares concentration of one variable with respect to another.
- If x and y are two random variables, then their covariance is

$$Cov(x, y) = E\{(x - E\{x\})(y - E\{y\})\}$$

- Note that $Cov(x, y) = Cov(y, x)$ and $Cov(x, x) = Var\{x\}$
- If x and y have 'similar' values, then $E\{x\} \approx E\{y\}$ and

$$Cov(x, y) \approx Var\{x\} \approx Var\{y\}$$

- If x and y are not 'similar', then $Cov(x, y) \approx 0$

CORRELATION

- The ratio of covariance with respect to variances of each variable is called correlation

$$Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var\{x\}Var\{y\}}}$$

CORRELATION

- The ratio of covariance with respect to variances of each variable is called correlation

$$Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var\{x\} Var\{y\}}}$$

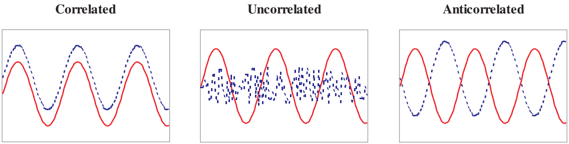
- If $x = y$, then $Corr(x, y) = 1$ (for $Cov(x, x) = Var\{x\}$)

CORRELATION

- The ratio of covariance with respect to variances of each variable is called correlation

$$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var\{x\}Var\{y\}}}$$

- If $x = y$, then $Corr(x,y) = 1$ (for $Cov(x,x) = Var\{x\}$)



$Corr(x,y) = 1$ $\qquad$ $Corr(x,y) = 0$ $\qquad$ $Corr(x,y) = -1$

CORRELATION MATRIX

- Correlations (or covariances) can tell us how ‘similar’ are two random variables.

	Age	Gender	Income	Expenses	H. owner	C. score
Age	1,0	0,6	0,9	0,6	0,4	0,5
Gender	0,6	1,0	0,2	1,0	-0,2	-0,3
Income	0,9	0,2	1,0	0,2	0,7	0,9
Expenses	0,6	1,0	0,2	1,0	-0,2	-0,3
H. owner	0,4	-0,2	0,7	-0,2	1,0	0,9
C. score	0,5	-0,3	0,9	-0,3	0,9	1,0

CORRELATION MATRIX

- Correlations (or covariances) can tell us how 'similar' are two random variables.
- Below is the correlation matrix showing correlations of each pair of variables in our database

	Age	Gender	Income	Expenses	H. owner	C. score
Age	1,0	0,6	0,9	0,6	0,4	0,5
Gender	0,6	1,0	0,2	1,0	-0,2	-0,3
Income	0,9	0,2	1,0	0,2	0,7	0,9
Expenses	0,6	1,0	0,2	1,0	-0,2	-0,3
H. owner	0,4	-0,2	0,7	-0,2	1,0	0,9
C. score	0,5	-0,3	0,9	-0,3	0,9	1,0

DECISION-MAKING UNDER UNCERTAINTY

- If the outcomes of decisions are not certain, then the knowledge of the utility function is not sufficient for making choices.

DECISION-MAKING UNDER UNCERTAINTY

- If the outcomes of decisions are not certain, then the knowledge of the utility function is not sufficient for making choices.
- If we know probabilities of utilities for different decisions, then we can 'predict' (estimate) the utilities and choose according to the best prediction.

DECISION-MAKING UNDER UNCERTAINTY

- If the outcomes of decisions are not certain, then the knowledge of the utility function is not sufficient for making choices.
- If we know probabilities of utilities for different decisions, then we can 'predict' (estimate) the utilities and choose according to the best prediction.
- There are many ways to estimate (predict) using probabilities. The most popular are using the expected value and the maximum likelihood (the most probable value).

MAXIMUM EXPECTED UTILITY PRINCIPLE

- In brief, choose the decision that yields the highest expectation.

MAXIMUM EXPECTED UTILITY PRINCIPLE

- In brief, choose the decision that yields the highest expectation.
- If U_1, U_2,...,U_n are the possible utility values (i.e. the outcomes), and $P(U_1)$, $P(U_2)$,...,$P(U_n)$ are their probabilities, then we can compute the expected utility

$$E\{u\} = U_1 P(U_1) + \cdots + U_n P(U_n)$$

MAXIMUM EXPECTED UTILITY PRINCIPLE

- In brief, choose the decision that yields the highest expectation.
- If U_1, U_2,...,U_n are the possible utility values (i.e. the outcomes), and $P(U_1)$, $P(U_2)$,...,$P(U_n)$ are their probabilities, then we can compute the expected utility

$$E\{u\} = U_1 P(U_1) + \cdots + U_n P(U_n)$$

- If d_1 and d_2 are two alternative decisions, then we choose d_1 if

$$E\{u \mid d_1\} \geq E\{u \mid d_2\}$$

MAXIMUM EU EXAMPLE

Example

- Lottery A, in which you can win $1000, but also you may loose $100;

MAXIMUM EU EXAMPLE

Example

- Lottery A, in which you can win $1000, but also you may loose $100;
- Lottery B, in which you can win $100, but you can loose $10.

MAXIMUM EU EXAMPLE

Example

- Lottery A, in which you can win $\$1000$, but also you may loose $\$100$;
- Lottery B, in which you can win $\$100$, but you can loose $\$10$.
- Suppose also that the probability of winning in both lotteries is $\frac{1}{2}$

MAXIMUM EU EXAMPLE

Example

- Lottery A, in which you can win $\$1000$, but also you may loose $\$100$;
- Lottery B, in which you can win $\$100$, but you can loose $\$10$.
- Suppose also that the probability of winning in both lotteries is $\frac{1}{2}$

$$E\{u \mid A\} = \$1000\frac{1}{2} - \$100\frac{1}{2} = \$450$$

$$E\{u \mid B\} = \$100\frac{1}{2} - \$10\frac{1}{2} = \$45$$

MAXIMUM EU EXAMPLE

Example

- Lottery A, in which you can win $\$1000$, but also you may loose $\$100$;
- Lottery B, in which you can win $\$100$, but you can loose $\$10$.
- Suppose also that the probability of winning in both lotteries is $\frac{1}{2}$

$$E\{u \mid A\} = \$1000\frac{1}{2} - \$100\frac{1}{2} = \$450$$

$$E\{u \mid B\} = \$100\frac{1}{2} - \$10\frac{1}{2} = \$45$$

Thus, we prefer A to B.

SUMMARY

- Probability allows us to understand our data better

SUMMARY

- Probability allows us to understand our data better
- The analysis of probabilities can show that some events or variables are not independent, and therefore the data is not entirely random (i.e. the uncertainty is not so great).

SUMMARY

- Probability allows us to understand our data better
- The analysis of probabilities can show that some events or variables are not independent, and therefore the data is not entirely random (i.e. the uncertainty is not so great).
- For businesses, the information from their corporate data can lead to a discovery of new knowledge.

SUMMARY

- Probability allows us to understand our data better
- The analysis of probabilities can show that some events or variables are not independent, and therefore the data is not entirely random (i.e. the uncertainty is not so great).
- For businesses, the information from their corporate data can lead to a discovery of new knowledge.
- This process is often called *knowledge discovery in databases*